

BEST AVAILABLE COPY

BIOTECH. LOGY
SYSTEMS
BRANCH



RAW SEQUENCE LISTING
ERROR REPORT

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer-readable form:

RECEIVED

Application Serial Number: 09/489,198
Source: 1646
Date Processed by STIC: 6-14-01

JUL 09 2001

TECH CENTER 1600/2900

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.

PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.

PATENTIN 2.1 e-mail help: patin21help@uspto.gov or phone 703-306-4119 (R. Wax)

PATENTIN 3.0 e-mail help: patin3help@uspto.gov or phone 703-306-4119 (R. Wax)

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 3.0 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW:

Checker Version 3.0

The Checker Version 3.0 application is a state-of-the-art Windows based software program employing a logical and intuitive user-interface to check whether a sequence listing is in compliance with format and content rules. Checker Version 3.0 works for sequence listings generated for the original version of 37 CFR §§1.821 – 1.825 effective October 1, 1990 (old rules) and the revised version (new rules) effective July 1, 1998 as well as World Intellectual Property Organization (WIPO) Standard ST.25.

Checker Version 3.0 replaces the previous DOS-based version of Checker, and is Y2K-compliant. Checker allows public users to check sequence listings in Computer Readable form (CRF) before submitting them to the United States Patent and Trademark Office (USPTO).

Use of Checker prior to filing the sequence listing is expected to result in fewer errored sequence listings, thus saving time and money.

Checker Version 3.0 can be down loaded from the USPTO website at the following address:

<http://www.uspto.gov/web/offices/pac/checker>

Raw Sequence Listing Error Summary

ERROR DETECTED

SUGGESTED CORRECTION

SERIAL NUMBER: 09/489,198

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1 Wrapped Nucleics
 Wrapped Aminos The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."

- 2 Invalid Line Length The rules require that a line not exceed 72 characters in length. This includes white spaces.

- 3 Misaligned Amino
 Numbering The numbering under each 5th amino acid is misaligned. Do not use tab codes between numbers; use space characters, instead.

- 4 Non-ASCII The submitted file was not saved in ASCII(DOS) text, as required by the Sequence Rules. Please ensure your subsequent submission is saved in ASCII text.

- 5 Variable Length Sequence(s) contain n's or Xaa's representing more than one residue. Per Sequence Rules, each n or Xaa can only represent a single residue. Please present the maximum number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.

- 6 PatentIn 2.0
 "bug" A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s) . Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.

- 7 Skipped Sequences
 (OLD RULES) Sequence(s) missing. If intentional, please insert the following lines for each skipped sequence:
 (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)
 (i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)
 This sequence is intentionally skipped

 Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to include the skipped sequences.

- 8 Skipped Sequences
 (NEW RULES) Sequence(s) missing. If intentional, please insert the following lines for each skipped sequence.
 <210> sequence id number
 <400> sequence id number
 000

- 9 ✓ Use of n's or Xaa's
 (NEW RULES) Use of n's and/or Xaa's have been detected in the Sequence Listing.
 Per 1.823 of Sequence Rules, use of <220>-<223> is MANDATORY if n's or Xaa's are present
 In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.

- 10 Invalid <213>
 Response Per 1.823 of Sequence Rules, the only valid <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is required when <213> response is Unknown or is Artificial Sequence

- 11 Use of <220> Sequence(s) missing the <220> "Feature" and associated numeric identifiers and responses. Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section.
 (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)

- 12 PatentIn 2.0
 "bug" Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.

#10
Cmen
1-9-10
1646

RAW SEQUENCE LISTING

DATE: 06/14/2001

PATENT APPLICATION: US/09/489,198

TIME: 10:13:12

Input Set : A:\Sequence.txt

Output Set: N:\CRF3\06142001\I489198.raw

3 <110> APPLICANT: CHUGAI SEIYAKU KABUSHIKI KAISHA.
 5 <120> TITLE OF INVENTION: Gene screening method using nuclear receptor
 7 <130> FILE REFERENCE: C1-901PCT
 C--> 9 <140> CURRENT APPLICATION NUMBER: US/09/489,198
 C--> 10 <141> CURRENT FILING DATE: 2000-01-20
 12 <150> PRIOR APPLICATION NUMBER: JP 09/212624
 W--> 13 <151> PRIOR FILING DATE: 1997-7-22 → 1997-07-22
 15 <160> NUMBER OF SEQ ID NOS: 4
 17 <170> SOFTWARE: PatentIn Ver. 2.0

Does Not Comply
 Corrected Diskette Needed
 See pp. 1, 2, 5

ERRORED SEQUENCES

19 <210> SEQ ID NO: 1
 20 <211> LENGTH: 507
 E--> 21 <212> TYPE: RPT → Valid <212> responses:
 22 <213> ORGANISM: Mus musculus
 24 <400> SEQUENCE: 1
 25 Met Thr Gln Ala Val Lys Leu Ala
 26 1 5
 27 Ser Arg Val Phe His Arg Ile His Leu Pro Leu Gln Leu Asp Ala Ser
 28 10 15 20
 29 Leu Gly Ser Arg Gly Ser Glu Ser Val Leu Arg Ser Leu Ser Asp Ile
 30 25 30 35 40
 31 Pro Gly Pro Ser Thr Leu Ser Phe Leu Ala Glu Leu Phe Cys Lys Gly
 32 45 50 55
 33 Gly Leu Ser Arg Leu His Glu Leu Gln Val His Gly Ala Ala Arg Tyr
 34 60 65 70
 35 Gly Pro Ile Trp Ser Gly Ser Phe Gly Thr Leu Arg Thr Val Tyr Val
 36 75 80 85
 37 Ala Asp Pro Thr Leu Val Glu Gln Leu Leu Arg Gln Glu Ser His Cys
 38 90 95 100
 39 Pro Glu Arg Cys Ser Phe Ser Ser Trp Ala Glu His Arg Arg Arg His
 40 105 110 115 120
 41 Gln Arg Ala Cys Gly Leu Leu Thr Ala Asp Gly Glu Glu Trp Gln Arg
 42 125 130 135
 43 Leu Arg Ser Leu Leu Ala Pro Leu Leu Leu Arg Pro Gln Ala Ala Ala
 44 140 145 150
 45 Gly Tyr Ala Gly Thr Leu Asp Asn Val Val Arg Asp Leu Val Arg Arg
 46 155 160 165
 47 Leu Arg Arg Gln Arg Gly Arg Gly Ser Gly Leu Pro Gly Leu Val Leu
 48 170 175 180
 49 Asp Val Ala Gly Glu Phe Tyr Lys Phe Gly Leu Glu Ser Ile Gly Ala
 50 185 190 195 200
 51 Val Leu Leu Gly Ser Arg Leu Gly Cys Leu Glu Ala Glu Val Pro Pro
 52 205 210 215
 53 Asp Thr Glu Thr Phe Ile His Ala Val Gly Ser Val Phe Val Ser Thr

RAW SEQUENCE LISTING

DATE: 06/14/2001

PATENT APPLICATION: US/09/489,198

TIME: 10:13:12

Input Set : A:\Sequence.txt

Output Set: N:\CRF3\06142001\I489198.raw

```

54          220          225          230
55 Leu Leu Thr Met Ala Met Pro Asn Trp Leu His His Leu Ile Pro Gly
56          235          240          245
57 Pro Trp Ala Arg Leu Cys Arg Asp Trp Asp Gln Met Phe Ala Phe Ala
58          250          255          260
59 Gln Arg His Val Glu Leu Arg Glu Gly Glu Ala Ala Met Arg Asn Gln
60 265          270          275          280
61 Gly Lys Pro Glu Glu Asp Met Pro Ser Gly His His Leu Thr His Phe
62          285          290          295
63 Leu Phe Arg Glu Lys Val Ser Val Gln Ser Ile Val Gly Asn Val Thr
64          300          305          310
65 Glu Leu Leu Leu Ala Gly Val Asp Thr Val Ser Asn Thr Leu Ser Trp
66          315          320          325
67 Thr Leu Tyr Glu Leu Ser Arg His Pro Asp Val Gln Thr Ala Leu His
68          330          335          340
69 Ser Glu Ile Thr Ala Gly Thr Arg Gly Ser Cys Ala His Pro His Gly
70 345          350          355          360
71 Thr Ala Leu Ser Gln Leu Pro Leu Leu Lys Ala Val Ile Lys Glu Val
72          365          370          375
73 Leu Arg Leu Tyr Pro Val Val Pro Gly Asn Ser Arg Val Pro Asp Arg
74          380          385          390
75 Asp Ile Arg Val Gly Asn Tyr Val Ile Pro Gln Asp Thr Leu Val Ser
76          395          400          405
77 Leu Cys His Tyr Ala Thr Ser Arg Asp Pro Thr Gln Phe Pro Asp Pro
78          410          415          420
79 Asn Ser Phe Asn Pro Ala Arg Trp Leu Gly Glu Gly Pro Thr Pro His
80 425          430          435          440
81 Pro Phe Ala Ser Leu Pro Phe Gly Phe Gly Lys Arg Ser Cys Ile Gly
82          445          450          455
83 Arg Arg Leu Ala Glu Leu Glu Leu Gln Met Ala Leu Ser Gln Ile Leu
84          460          465          470
85 Thr His Phe Glu Val Leu Pro Glu Pro Gly Ala Leu Pro Ile Lys Pro
86          475          480          485
87 Met Thr Arg Thr Val Leu Val Pro Glu Arg Ser Ile Asn Leu Gln Phe
88          490          495          500
89 Val Asp Arg
90 505

```

92 <210> SEQ ID NO: 2

93 <211> LENGTH: 508

E--> 94 <212> TYPE: RPT

95 <213> ORGANISM: Homo sapiens

97 <400> SEQUENCE: 2

98 Met Thr Gln Thr Leu Lys Tyr Ala Ser Arg Val Phe His Arg Val Arg

99 1

5

10

15

100 Trp Ala Pro Glu Leu Gly Ala Ser Leu Gly Tyr Arg Glu Tyr His Ser

101 20

25

30

102 Ala Arg Arg Ser Leu Ala Asp Ile Pro Gly Pro Ser Thr Pro Ser Phe

103 35

40

45

104 Leu Ala Glu Leu Phe Cys Lys Gly Gly Leu Ser Arg Leu His Glu Leu

valid responses:

- DNA

- RNA

- PRT

RAW SEQUENCE LISTING

DATE: 06/14/2001

PATENT APPLICATION: US/09/489,198

TIME: 10:13:12

Input Set : A:\Sequence.txt

Output Set: N:\CRF3\06142001\I489198.raw

```

105      50      55      60
106 Gln Val Gln Gly Ala Ala His Phe Gly Pro Val Trp Leu Ala Ser Phe
107 65      70      75      80
108 Gly Thr Val Arg Thr Val Tyr Val Ala Ala Pro Ala Leu Val Glu Glu
109      85      90      95
110 Leu Leu Arg Gln Glu Gly Pro Arg Pro Glu Arg Cys Ser Phe Ser Pro
111      100      105      110
112 Trp Thr Glu His Arg Arg Cys Arg Gln Arg Ala Cys Gly Leu Leu Thr
113      115      120      125
114 Ala Glu Gly Glu Glu Trp Gln Arg Leu Arg Ser Leu Leu Ala Pro Leu
115      130      135      140
116 Leu Leu Arg Pro Gln Ala Ala Ala Arg Tyr Ala Gly Thr Leu Asn Asn
117 145      150      155      160
118 Val Val Cys Asp Leu Val Arg Arg Leu Arg Arg Gln Arg Gly Arg Gly
119      165      170      175
120 Thr Gly Pro Pro Ala Leu Val Arg Asp Val Ala Gly Glu Phe Tyr Lys
121      180      185      190
122 Phe Gly Leu Glu Gly Ile Ala Ala Val Leu Leu Gly Ser Arg Leu Gly
123      195      200      205
124 Cys Leu Glu Ala Gln Val Pro Pro Asp Thr Glu Thr Phe Ile Arg Ala
125      210      215      220
126 Val Gly Ser Val Phe Val Ser Thr Leu Leu Thr Met Ala Met Pro His
127 225      230      235      240
128 Trp Leu Arg His Leu Val Pro Gly Pro Trp Gly Arg Leu Cys Arg Asp
129      245      250      255
130 Trp Asp Gln Met Phe Ala Phe Ala Gln Arg His Val Glu Arg Arg Glu
131      260      265      270
132 Ala Glu Ala Ala Met Arg Asn Gly Gly Gln Pro Glu Lys Asp Leu Glu
133      275      280      285
134 Ser Gly Ala His Leu Thr His Phe Leu Phe Arg Glu Glu Leu Pro Ala
135      290      295      300
136 Gln Ser Ile Leu Gly Asn Val Thr Glu Leu Leu Leu Ala Gly Val Asp
137 305      310      315      320
138 Thr Val Ser Asn Thr Leu Ser Trp Ala Leu Tyr Glu Leu Ser Arg His
139      325      330      335
140 Pro Glu Val Gln Thr Ala Leu His Ser Glu Ile Thr Ala Ala Leu Ser
141      340      345      350
142 Pro Gly Ser Ser Ala Tyr Pro Ser Ala Thr Val Leu Ser Gln Leu Pro
143      355      360      365
144 Leu Leu Lys Ala Val Val Lys Glu Val Leu Arg Leu Tyr Pro Val Val
145      370      375      380
146 Pro Gly Asn Ser Arg Val Pro Asp Lys Asp Ile His Val Gly Asp Tyr
147 385      390      395      400
148 Ile Ile Pro Lys Asn Thr Leu Val Thr Leu Cys His Tyr Ala Thr Ser
149      405      410      415
150 Arg Asp Pro Ala Gln Phe Pro Glu Pro Asn Ser Phe Arg Pro Ala Arg
151      420      425      430
152 Trp Leu Gly Glu Gly Pro Thr Pro His Pro Phe Ala Ser Leu Pro Phe
153      435      440      445

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/489,198

DATE: 06/14/2001

TIME: 10:13:12

Input Set : A:\Sequence.txt

Output Set: N:\CRF3\06142001\I489198.raw

```
154 Gly Phe Gly Lys Arg Ser Cys Met Gly Arg Arg Leu Ala Glu Leu Glu
155      450                      455                      460
156 Leu Gln Met Ala Leu Ala Gln Ile Leu Thr His Phe Glu Val Gln Pro
157 465                      470                      475                      480
158 Glu Pro Gly Ala Ala Pro Val Arg Pro Lys Thr Arg Thr Val Leu Val
159                      485                      490                      495
160 Pro Glu Arg Ser Ile Asn Leu Gln Phe Leu Asp Arg
161                      500                      505
```

9/489, 198

p. 5

<210> 4 Seq. # 4
<211> 2362
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)... (1524)

There is an "n" at position 1926
in sequence # 4. It is mandatory
to have <220> to <223> features to
explain what the "n" represents.
See # 9 on the Error Summary
Sheet.

VERIFICATION SUMMARY

DATE: 06/14/2001

PATENT APPLICATION: US/09/489,198

TIME: 10:13:13

Input Set : A:\Sequence.txt

Output Set: N:\CRF3\06142001\I489198.raw

L:9 M:270 C: Current Application Number differs, Replaced Application Number
L:10 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:13 M:256 W: Invalid Numeric Header Field, Wrong Prior FILING DATE:YYYY-MM-DD
L:21 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:
L:94 M:310 E: (3) Wrong or Missing Sequence Type, TYPE:
L:399 M:258 W: Mandatory Feature missing, <223> not found for SEQ ID#:4
L:399 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4